DOCUMENT RESUME

ED 189 057 SP 016 241

AUTHOR Borich, Gary D.
TITLE Three Models for Conducting Follow-Up Studies of Teacher Education and Training.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY Organisation for Economic Cooperation and Development, Paris (France).
NOTE 35p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Gains: *Evaluation Criteria: *Evaluation Methods: *Followup Studies: *Inservice Teacher Education: Needs Assessment: *Preservice Teacher Education: *Program Implementation: Program Improvement: Teacher Effectiveness

ABSTRACT

This paper presents three methodological models that can be employed in conducting follow-up studies of preservice and inservice teacher training. The needs assessment model evaluates effectiveness by the discrepancy between "what is" and "what should be." The relative gain model assesses relative pupil improvement. The process-product model of evaluation tests the appropriateness of behaviors taught to teachers and to what degree the program can produce these behaviors. Implementation steps for each model are included, along with a section of "case studies" that show the different applications possible with each model. (CJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Three Models for Conducting Follow-up Studies Of Teacher Education and Training*

Gary D. Borich

The University of Texas at Austin

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Carol Walton
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED189057

Teaching competencies derived from process-product studies (Borich, 1979; Borich and Kash, 1979; Medley, 1978; Rosenshine, 1975) have contributed in some degree to the content of inservice evaluations of teacher training. However, because the intent of these studies has been limited to investigating teaching behavior vis-a-vis pupil outcomes, they have been less useful in suggesting the methodology with which institutions could conduct inservice evaluations of their training programs. This paper presents three methodological models that can be employed in conducting follow-up studies of preservice and inservice training. These models do not represent the only methodologies available for this purpose but do illustrate the variety of methods that can be employed by a training institution. As in any field-study, the "best" methodology is dictated by many factors, not the least of which will be the resources at hand, time and commitment of those conducting the study, the requirements and policies providing the initial impetus, and the objectives of the training institution.

The three models to be presented are arranged in an order that reflects the time and expense generally required for their implemementation, with the least costly and least time-consuming appearing first. For

SP 016241

*This paper was prepared for the Organization for Economic Cooperation and Development (OECD) and represents a portion of the author's final report as a member of OECD Expert's Committee on the Evaluation of Inservice Education and Training. See also the Journal of Teacher Education, 1979, Jan.-Feb. XXX, 1, p. 77-86, for a related portion of the author's report.

purposes of this paper, the three models have been titled Needs Assessment, Relative Gain, and Process-Product. However, they have no commonly accepted titles and are intended as generic or theoretical models from which other, more context-specific versions can be derived. Thus, this paper might best be described as a concept paper from which readers can derive ideas for composing new models appropriate to their own training objectives, resources and institutions.

NEEDS ASSESSMENT MODEL

Overview

Training institutions are continually in search of ways to improve their training programs. Methods for determining needed areas of improvement have ranged from telephone surveys to full-fledged research studies complete with control and experimental groups. Perhaps most used among these methods has been the follow-up questionnaire mailed to recent graduates for the purpose of eliciting their opinions about the training they received and the extent to which this training may or may not have prepared them to meet the immediate demands of their teaching environment. While many of these surveys are creatively conceived and exhaustively carried out, their ultimate utility rests on the extent to which they provide direct and unambiguous recommendations for program improvement. Unfortunately, the excitement and challange of the survey process can turn to disappointment when evaluators are faced with the task of making unambiguous recommendations from the data they receive.

One approach to conducting a follow-up study is to design the survey instrument in such a way as to "lock in" from the outset the type and

JUL 3 1980

quality of the data that will be received. This can be accomplished by employing a specific model for collecting the data which establishes prior to questionnaire development the precise scheme by which the data will be analyzed and interpreted. This can be done by designing the questionnaire in such a manner that respondents provide data in a form that can be weighted and prioritized so that responses are linked to a practical decision framework for program improvement. Conceptualizing possible program weaknesses as relative "needs" and utilizing a needs assessment model for data collection is one means of accomplishing this.

A training need can be defined as a discrepancy between an educational goal and trainee performance in relation to this goal. The process of identifying training needs can be conceptualized as a discrepancy analysis that identifies the two polar positions of "what is" and "what should be." Training programs can apply this model by defining "what is" as the measured behaviors, skills and competencies of the trainee and "what should be" as the goals or intents of the training program. The discrepancy (or distance) between these two poles can then be used as an index of the effectiveness of the training program in reaching its intended goals and to identify components of the training process that fail to engender specified behaviors. Discrepancies can be prioritized by a panel of trainers or by statistical techniques that weight the relative importance of each goal statement from values assigned to them by the respondents. Prioritized discrepancies, ranked in descending order, provide the framework for deciding what parts of the program to modify or revise.

Model

Following are the steps involved in implementing the needs assessment model.

1. List competencies. Competency statements can be derived from the competency implications of teacher effectiveness studies or from the intents and objectives of teacher trainers, or both. Inservice teachers, trainees and program administrators can assist in this task by supplying competency statements derived from a study of the activities and materials used in the training program or from a list of program objectives. All competency statements are checked against program activities and materials to insure that they actually represent program objectives. These competency statements are then used in constructing the survey instrument.

2. Survey inservice teachers. All or a sample of the trainees who have completed training are asked to rate (a) the relevance of each competency to their current job function (or perceived future job function) and (b) their current level of attainment of each competency. A typical questionnaire might take the following two-part format.

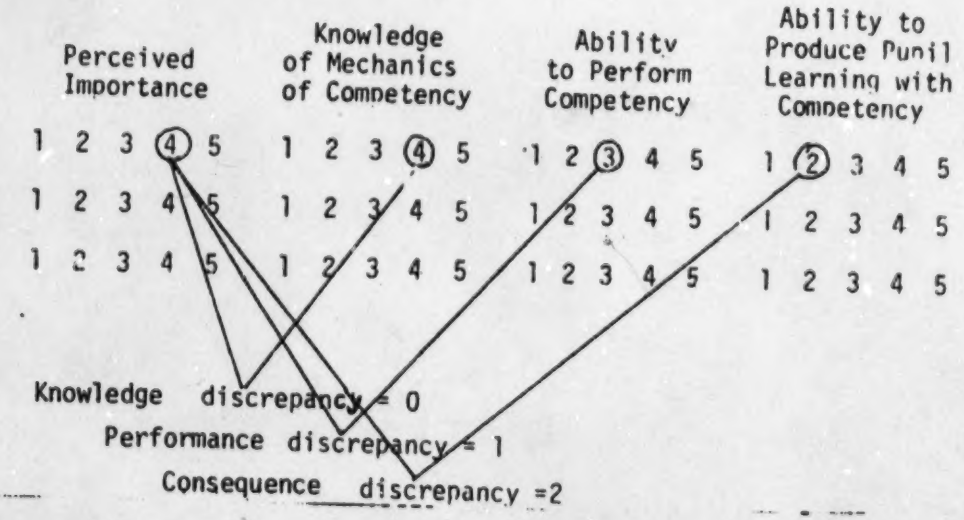
Competency	Perceived Relevance					Perceived Level of Attainment				
	Low				High	Low				High
1.	1	2	3	4	5	1	2	3	4	5
2.	1	2	3	4	5	1	2	3	4	5
3.	1	2	3	4	5	1	2	3	4	5
4.	1	2	3	4	5	1	2	3	4	5

An alternative format for rating competency attainment is to divide each competency statement into "knowledge" competence, "performance" competence and "consequence" competence. These terms can be defined on the questionnaire in the following manner.

> Knowledge competence: Ability to accurately recall, paraphrase, or summarize the procedural mechanics of the behavior on a paper and pencil test.
>
> Performance competence: Ability to accurately execute the behavior in a real or simulated environment in the presence of an observer.

Consequence competence: Ability to elicit learning from pupils (as recorded on tests of affective and/or cognitive achievement) by using the behavior in the classroom.

These distinctions require the teacher to make finer judgments in rating each competency and in turn permit a more refined evaluation of the training program. A questionnaire incorporating these competency dimensions might take the following form:



Each competency then yields for each respondent three discrepancy scores which indicate the effectiveness of the training program in producing (a) trainee knowledge, (b) trainee performance, and (c) pupil consequences. Using these discrepancy scores as a guide, components of the training program can be revised to produce increased "knowledge," "performance," and/or "pupil impact." The three discrepancy scores above might indicate that field experiences for this competency (performance and consequence) need improvement but classroom instruction (knowledge) is adequate.

3. Rank Competencies. Competencies are then ranked according to ratings obtained on the questionnaire. For each competency the difference between perceived importance and perceived level of attainment is calculated across the three dimensions: knowledge, performance, and consequence. These differences are ordered according to magnitude or relative weight, calculated by multiplying the discrepancy score by the average perceived importance determined over all respondents. In the above example, if the average perceived importance of competency 1 were 2.5, the resulting knowledge discrepancy would be 0.0, the resulting performance discrepancy would be 2.5 and the resulting consequence discrepancy would be 5.0. Other competencies deemed either more or less important than this competency would have their discrepancies adjusted accordingly. This weighted ordering takes into account that a small discrepancy on one competency may be of greater perceived importance than a large discrepancy on another competency. Those discrepancies with the greatest positive rank difference would have the highest priority for revising the training program.

4. Compare High Priority Competencies with the Content of the Training Program. High priority competencies determined form the above analysis are compared to the instructional experiences, components, and materials that receive high priority in the training curriculum. The instructional staff might examine instructional time devoted to the competency, the clarity of the instruction, adequacy of the training materials, and the number of minutes or hours alotted to students for practicing the competency in order to determine the emphasis that the training program is actually placing on the competency. When a competency is highly valued but poorly performed, the problem may derive from insufficient rather than ineffective training.

5. Revise Program or Revise Competency. Where possible, the emphasis of the training program is modified to match high priority competencies. Or, if altering the training program to emphasize a particular high priority competency is not cost effective, other training resources (e.g. self-paced modules, programmed texts) or other alternatives (e.g., agencies and institutions at which the inservice teacher may obtain the needed training) are recommended to program graduates.

Further Considerations

The needs assessment model can be extended and adapted to meet a variety of institutional needs. For example, the needs assessment instrument could be used in conjunction with a similar survey completed by supervisors or administrators in order to corroborate the subjective responses of the teachers. An evaluation of training, for instance, might be based on the mean discrepancy across teachers and supervisors, thereby taking into account a second and presumably more objective group of respondents. Or, competencies for which the reported level of attainment differs dramatically from supervisor to teacher can be withheld from analysis pending clarification from other data sources, such as the classroom observation of teachers.

Evaluations employing the needs assessment model can have multiple purposes. These purposes derive form the nature of the needs data which can be employed with equal effectiveness for making either formative or summative judgments about the training program. Formative data revealing the perceived importance of the competencies taught can serve as a check on the relevance of the training and as a guide to what additional training

may be needed. Summative data revealing the level to which trainees attained the competencies compared with trainees from the other programs or institutions can serve as an overall check on the program. The versatility of these data make the needs assessment model less restrictive and more developmental than other approaches to the evaluation of training.

The needs assessment model is essentially a self-evaluative procedure which relies on the judgments of teachers about their own performance. The assumption underlying the needs model is that the performer (teacher) can best judge his or her own performance and, when explicitly asked to do so, can make an objective judgment. This assumption is most tenable when the purpose of the data collection is the evaluation of training and not the evaluation of individual teachers. If desired, efforts to make the needs assessment model more objective can include supervisor-administrator ratings or limited follow-up visitation. These additions can enhance the credibility of the self-reports and provide an additional vantage point from which to judge discrepancies between program intents and the performance of trainees.

An important practical characteristic of the needs assessment model is the ease with which it can be implemented. It is sufficiently direct that data analysis and instrument construction are no more complex than with any type of follow-up survey; yet, the amount and interpretability of the data it yields is considerably greater than many types of follow-up questionnaires. Consequently, it is a model that is easily implementable by teacher trainers who need immediate feedback on the effectiveness of program experiences and materials and who have limited resources.

Finally, we must note the definition of evaluation implied by the needs assessment model: determining the congruence between what "should be" and "what is," i.e., between what the teacher should be able to do and what the teacher can do. The evaluation is complete when the training program has objectively determined the discrepancy between these two poles. This definition calls for the development of goals and objectives (in the form of competency statements) and an assessment of whether these goals and objectives have been met. This is accomplished by obtaining self-report data about both the perceived value of the training objectives and the level to which trainees have attained the objectives.

RELATIVE GAIN MODEL

Overview

Inclusion of the relative gain model among the models considered in this paper stems from the belief among some that teachers, schools, and educational programs should be directly accountable to their constituencies. This belief has prompted a number of states to pass laws or policies making evaluation and accountability procedures mandatory at the school and school district level (Gage, 1973). In response to this mandate, procedures have been developed to compare student performance in different classrooms in an effort to establish minimum standards of pupil growth for which all teachers could be held accountable.

The primary assumption underlying the relative gain model is that an "effective" training program emphasizes the objective assessment of its trainees. Some proponents of educational accountability have argued that the needs assessment model is too susceptible to individual bias to

provide accurate data upon which to base accountability decisions and that a more valid index of teacher effectiveness is pupil achievement. The logic underlying this position reduces to the notion that if a teacher is doing his or her job well, that teacher's pupils should exhibit more learning than those taught by a teacher who is not doing his or her job well. This logic can be applied to the evaluation of training, however, only if (1) those factors over which the teacher has no control can be identified and separated from the evaluation of the training program, and (2) the phrase "doing his or her job well" can be translated into meaningful units of pupil achievement. To resolve the first problem, pupil achievement scores can be adjusted to account for differences among pupils prior to instruction. To resolve the second problem, traditionally the more difficult, an effort can be made to make pupil tests as sensitive as possible to the competencies stressed in the training program.

By focusing on pupil performance, the relative gain model measures behavior at least one step removed form the training program. The effect of training must register not only on measures of teacher behavior but also on tests of pupil behavior. The model rests on the assumption that teacher competencies can be translated directly into pupil competencies

and that potentially confounding variables can be statistically controlled to an extent which allows the effects of teacher training to filter down and be measured in units of pupil performance.

There are two procedures that can be used to increase the likelihood of obtaining valid measures of relative gain. The first involves the use of criterion-referenced rather than norm referenced tests to align training with testing. Nationally normed tests provide only a single score on very general objectives and their content is sometimes only partially relevant to the specific objectives of a particular training program. Also, norm-referenced tests are usually administered only once a year on a prespecified date which may not follow training. For this reason, criterion-referenced tests of pupil performance are recommended for the relative gain model. These tests are designed to measure only pupil outcomes that are related to the content areas for which training has been provided and are usually constructed by program personnel who can best control the time of their administration.

Another procedure that strenghtens the link between teacher and pupil behavior is the clustering of competencies. Here, competencies are grouped according to their expected effects upon pupil behavior. Criterion-referenced tests are constructed to measure the joint or interactive effect of a competency cluster, thereby permitting a larger chunk of training to be evaluated at one time. While the attainment of a single competency may engender little, if any, change in pupil growth, the attainment of several closely related behaviors may substantially heighten the pupil effects that can be logically expected and measured. Both competency clustering and criterion-referenced testing are essential elements of the relative gain model.

Model

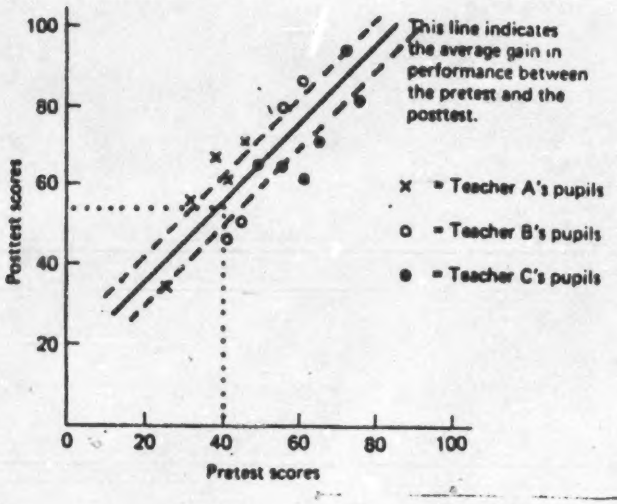
Following are the steps used in implementing the relative gain model.

1. List Competencies. As in the case of the needs assessment model, implementation begins with identification of the competencies to be measured. These can be derived directly from a study of the materials and activities within the training program or from a list of program objectives. The latter source is considered less reliable since the program may not be teaching all that it is said to teach. But, the former source may require considerably more time and expense, particularly if it involves direct observation and study of program materials. Competencies derived from process-product studies, especially those that have been shown to relate positively and significantly to pupil outcomes, are usually given highest priority in the relative gain model.

2. Construct Measures of Pupil Performance. Because this model makes no provision for measuring teacher behavior, the validity of all outcome data necessarily rests on the assumed relationship between the content the trainee is taught and the knowledge his or her pupils acquire. Hence, the importance of Step 1 to the success of the relative gain model cannot be overemphasized. It is from the competency lists compiled in Step 1 that post measures of pupil performance are derived. These measures usually take the form of criterion-referenced tests that assess pupil behaviors logically related to the competencies taught. Their construction requires considerable sensitivity on the part of the test developer in deciding what pupil outcomes can logically be expected from what teacher competencies. The validity of this procedure depends both on the clarity of the competencies and their capacity to engender specific pupil outcomes that cannot be influenced by factors unrelated to the training program.

3. Collect Pre and Post Pupil Data. Pretesting is an indispensible component of the relative gain model. It is the program's hedge against the criticism that pupil performance is, in part, determined by factors which the teacher cannot control. The pretest performance of pupils in the form of achievement and aptitude scores represents the covariable data with which posttest achievement must be statistically adjusted if the relative gain model is to measure pupil effects that are truly program specific. While pretest achievement data obtained on criterion-referenced tests can be the only variable used, the model is sufficiently general to accomodate any number of covariables, including pupil attitudes, interests, and previous experiences, that might otherwise confound the measurement and interpretation of teacher effects. To the extent that the evaluation ignores pupil variables that correlate with posttest pupil achievement, the accuracy and appropriateness of the relative gain model can be called into question. Pre and post achievement tests are used to assess each pupil's performance on all or a large sample of the relevant competencies within a given grade or age level. Generally, different sets of items are used for pre- and posttests to minimize the reactive effects of testing on pupil learning.

4. Plot Teacher Effects. Statistical analysis for the relative gain model begins with the construction of a prediction or regression equation between pre- and posttest pupil performance. This equation may utilize any number of covariables including pupil achievement and aptitude variables. For illustrative purposes, this discussion will be limited to the simplest case, that of pretest pupil achievement regressed on posttest achievement. First, pretest and posttest scores are plotted for each pupil within a given grade or age level. The pupils' teacher is also identified in this process. The scores of five pupils in each of three classes have been plotted below to illustrate the procedure. Teacher A and B have participated in the training program while Teacher C has not.



A regression line representing the average or typical relationship between pretest and posttest is drawn among these points on the plot. A standard linear prediction equation is used for this purpose. The program uses this equation to determine the typical posttest score for each pupil's pretest score. This is indicated by the dotted line (....) for a given pretest value. A typical equation might take the form:

Pupil's score on CRT at end of lesson = 1.4 (CRT score at beginning of lesson) + 1.75

If, for example, a pupil's score at the beginning of the lesson was 26 points, then the expected score at the end of the lesson would be 38.15, i.e., 1.4 (26) + 1.75 = 38.15. The regression equation describes the best relationship between input and outcome measures. It is positioned through the data points in such a way that the smallest error will be made in predicting each posttest score from its corresponding pretest score.

This general approach can be expanded to accommodate more than a single index of pupil entry behavior. When this is done, the procedure is called "multiple regression." The number of pre-measures that may be employed is determined by the combination of measures that are correlated with the posttest but not with each other. Generally, only a few pre-measures will meet this criterion. That is, only a few will contain unique information not supplied by other variables already entered into the equation.

5. Construct Confidence Band. When the actual outcome score for a given pupil is greater than the score predicted, performance is said to be "above" expectancy. Similarly, if the posttest score is lower than predicted, the pupil is said to be "below" expectancy. Above and below expectancy are relative determinations, since the standard for "above" and "below" is derived from a comparison of each pupil's actual pre- to posttest improvement with the average improvement of all pupils. Hence, the name, "relative gain model." Just how much an actual score must deviate from the expected score in order for it to be classified "above" or "below" is determined by constructing a confidence interval around expected scores. This band is illustrated above by the broken line on either side of the regression line. This band is a function of the standard error of estimate of the regression equation and its width can vary. One procedure is to use a band that is wide enough to accommodate approximately 2/3 of the pupils.

6. Construct Summary Table. Data from the relative gain model are reported in a table which takes the following form:

	Teacher A	*Teacher B*	*Teacher C*
Average pretest score	35	50	65
Average posttest score	57	65	73
Expected posttest score	50	65	80
Difference between expected and actual posttest scores	7	0	7
Percent of pupils who are:			
Above expectancy	60%	40%	20%
At expectancy	0	10%	0
Below expectancy	20%	40%	80%

Data are summarized by teacher and are averaged over all pupils for each class. Rows in the table indicate average pre, post, and expected post data and, most important, the percentage of pupils who scored above, below, or at expectancy. It is this latter set of data that is used to evaluate the training program. Since the criterion-referenced instrument was specifically constructed to match the objectives of the training program, most of the pupils of teachers who received instruction are expected to fall at or above expectancy, and most of the pupils of teachers who did not receive instruction are expected to fall below expectancy. There will be exceptions in either direction, but the average performance for each class should show the above trend if the training program was effective in teaching the intended competencies and if the teachers employed these competencies in the classroom. If this trend were reversed or if no discernable trend were apparent, the effectiveness of the training program, the appropriateness of the posttest measure, and/or the selection of entry level measures could be called into question. There is no substitute in the relative gain model for diligence in ruling out the latter two possibilities.

Further Considerations

The major advantages of the relative gain model are its focus on pupil performance, and its consideration of all pupil entry behaviors which might confound a test of training effectiveness. The model is limited, however, by its need for pupil performance tests in two alternative forms, its inability to compare teachers except on a grade-by-grade or subject-by-subject basis, and its requirement that classrooms evaluated contain approximately 20 or more students in order to arrive at stable estimates of expected pupil performance. Thus, the relative gain model, while serving as a stringent accountability measure requires considerably more time and expertise to implement than the needs assessment model. This latter point raises two questions: how often should the relative gain model be implemented, and on what unit of analysis--classroom, school, or school district--should it be based?

In response to the first question, the relative gain model can be used continuously to check the effectiveness of a training program. As pupils change from year to year along dimensions related to their achievement, so will the slope of the regression line on which the relative gain model is based. And, as this slope changes, the proportion of pupils falling above, at, and below expectancy will change. Each set of teachers exposed to the training program return to different schools, different pupils, and perhaps vastly different school districts. These variables may be sufficiently related to pre and post measures of pupil achievement to show the program in a different light. If program effectiveness varies considerably across groups of trainees, the characteristics on which the groups vary, e.g., pupil SES, school district, subject area, etc., should be included in subsequent evaluations as covariables in the regression equation. This flexibility is an appealing characteristic of the relative gain model; it can accommodate new information about pupils, teachers, or schools which might otherwise confound an evaluation.

In response to the second question, the relative gain model can be based on any of three units of analysis--classroom, school, school district--provided a sufficient number of classrooms, schools, or school districts are included in the study. If training is limited to a single school with only a few teachers participating, the relative gain model requires that classrooms contain at least 20 pupils each and results be generalized only to teachers within that specific grade in that school. When training is presented across schools, classroom averages can be employed. For dissemination and training projects within school districts, pupil performance can be aggregated by school, and for national projects where training materials are widely disseminated, it can be aggregated by school district. It is important in larger studies that the potentially confounding characteristics

of schools and school districts be quantified and added to the regression equation. Finally, separate equations should be constructed for each posttest measure, e.g., affective and cognitive, for each grade level at which the program is tested.

The relative gain model reflects the spirit of the accountability movement to a greater extent than the needs assessment model. Perhaps for this reason, the relative gain model may be somewhat anxiety provoking. This feeling is exacerbated by the fact that the teacher plays no direct role in study design, implementation, or data collection. In the needs assessment model, teacher opinion is elicited at several points, but the relative gain model bypasses the teacher to focus directly on pupil performance. This need not necessarily occur, however, as teachers can play an important role in helping to identify covariables that should be included in the regression analysis and, hence, "controlled" to insure that pupil outcomes reflect only the activity of the teacher. Also, teachers can be reminded that classroom data are pooled in order to make decisions about the training program. Although the teacher must be identified in this process, it is not the individual teacher on whom decisions are made, but rather the entire group of teachers who have received training.

The characteristic which perhaps most distinguishes the need assessment and relative gain models is the definition of evaluation implied by each. In the needs assessment model, evaluation is the discrepancy between "what is" and "what should be," with the teacher and/or supervisor providing data from which to determine the congruence between these two positions. In the relative gain model, evaluation is the assessment of normative improvement in pupil performance. By "normative improvement" is meant relative pupil improvement. Whether a particular pupil's score (or class mean) is "above" or "below"

expectancy depends not on the absolute value of that score but on its value relative to the average improvement of all pupils (classes). By definition, pupils' scores fall both above and below the regression line. Hence, there is always a forced distribution or relative ranking of scores. This ranking will occur regardless of how high or low pupils score on the pretest or how little variability exists among scores. The relative gain model does not test the effectiveness of training vis-a-vis some external criteria, but rather its ability to discriminate teachers who were trained from those who were not. Hence, the data reflect gains made by pupils of one group of teachers relative to those made by pupils of another group of teachers.

PROCESS-PRODUCT MODEL

Overview

The process-product model may be characterized as a hybrid model, reflecting in some respects the domain of evaluation and in other respects the domain of research. Since it spans these two domains, the process-product model has the capacity to serve both the teacher trainer and the researcher. This versatility does not come without a price, that price being the requirement that both the process behavior of the teacher and the performance behavior of pupils be measured.

Whether the process-product model serves the teacher trainer or the researcher depends on the way in which its data are used. Functioning as a tool for the evaluation of training, process-product relationships are used to test the appropriateness of the behaviors taught in training and to ascertain the degree to which the program can produce these behaviors in teachers. As in the relative gain model, "trained" and "untrained" groups of teachers can be used to substantiate the expectation that training increases the teacher's use of target behaviors. Teachers are also observed prior to and after training for the same purpose. The information obtained can be

used to gauge the importance of the behaviors emphasized in training (i.e., do they relate to pupil outcomes?) and also to determine the effectiveness of the training program in engendering these behaviors (i.e., does the trained teacher exhibit them more frequently than the untrained teacher?).

This information may be valuable not only to the trainer but also to the researcher and the program developer who is interested in program revision. Traditionally, the researcher observes teacher process behaviors with classroom observation instruments, usually incorporating a substantial number of teacher variables. The result is a large matrix of correlations depicting relationships between teacher process and pupil outcome variables. Since little or no theory may have been involved in selecting the relationships that were measured, it is not uncommon to find a substantial number of teacher behaviors in these studies exhibiting low frequencies, indicating that teachers had either little opportunity or no desire to use them on the occasions they were observed. Without theory, the results of process-product studies can be illusive, and the discovery of teacher-pupil relationships can be attributed as much to fortuitous probing as to systematic and informed investigation. For this reason the results of some process-product studies have been likened to the effects of a shotgun fired at long range: most of the shot misses the target, some barely misses it, and a little hits the bullseye.

The process-product model employed in the context of a training program has the distinct advantage of having an identifiable target. The objectives and rationale of the program provide the framework in which process and, therefore product, behaviors are measured and correlated. While some variation is anticipated, the behaviors taught can be expected to match the objectives of the teacher and, thus, to occur with sufficient frequency during observation to provide stable variance estimates. Thus, the shotgun

is exchanged for a rifle and the probability of a "hit" increased. While most process-product research has been program-free, the efficiency of the model improves considerably when it is used in the context of a specific training program with specific objectives. In this manner, its evaluative and research functions are combined.

Model

Following are the steps used in implementing the process-product model.

1. List Competencies. As in the two previous models, competencies must be identified from an examination of the training program or the objectives upon which the program is based. When findings from process-product studies provide the impetus for a training program, subsequent evaluation of the program can be considered an attempt at replicating the findings of the original studies in the context of a specific program. Even when objectives of the training program are not explicitly derived from earlier process-product findings, the competencies taught may be operationalized in the same manner as in these process-product investigations to allow a comparison of the two sets of results. Deriving competencies from a direct examination of program activities and materials is likely to be more time consuming but less risky: it increases the likelihood that behaviors taught but unspecified will be included in the evaluation and behaviors specified but clearly not taught will be eliminated.

2. Construct or Select Teacher and Pupil Measures. Developing instrumentation for the process-product model requires considerable circumspection and sensitivity. It is the most vulnerable link in the chain that connects the content of the training program and the performance of pupils. And, as noted earlier, this connection is moderated by the trainee who must not only learn the content of the training program but also implement it sufficiently to affect pupil performance. In addition, the model must take into consideration those variables beyond the teacher's control which can weaken the impact of training. Thus, the instrumentation for measuring teacher behavior taught by the program and pupil behavior taught by the trainee must be sensitive enough to record the effect of the program on pupils.

   Pupil outcomes should be derived directly from teacher behaviors which, in turn, should be derived from the goals and objectives of the training program. Instrumentation should reflect a tight, overlapping relationship among program goals, teacher behaviors, and pupil outcomes. This congruence is especially critical if the process-product model is to successfully focus the impact of the training program on the performance of pupils.

Measures for assessing teacher behavior are usually selected from a class of instruments called "classroom observation scales." These instruments focus the observer's attention on either low-inference (i.e., discrete and specific) or high-inference (i.e., general and cumulative) behavior. The choice of scale type and item content should be determined by the nature of the behavior being measured. Low-inference category systems are generally most appropriate when specific, discrete and context-related behaviors are taught by the training program, while high-inference systems are preferable when general, cumulative, and context-free behaviors are taught.

Instrumentation for recording pupil outcomes can be either standardized or, preferably, criterion-referenced. The link between teacher behavior and pupil outcome is as important as that between program content and teacher behavior. An adequate match between teacher behavior and pupil outcome can be achieved only if the instrument that records pupil performance is tailored to the explicit objectives of the teacher. Criterion-referenced tests that are relatively brief and highly focused on program content fulfill this requirement. They can be prepared for each segment of the training program, and alternate forms constructed for pre- and posttesting.

3. Observe In Situ. Systematic classroom observation is the characteristic that distinguishes the process-product model from the two previous models. The term "systematic" implies the rigorous application by two or more observers of classroom observation systems in each teacher's classroom over randomly selected occasions. One area of concern in this regard is the consistency or agreement (i.e. reliability) between two independent observations recorded on the same coding instrument. In order to determine that the behavior in question can be observed and recorded with some precision, the reliability of the coding system must be established by correlating observations recorded by different raters using the same instrument and observing a teacher for the same period of time. Thus, for at least part of the classroom observation, two or more coders must be used.

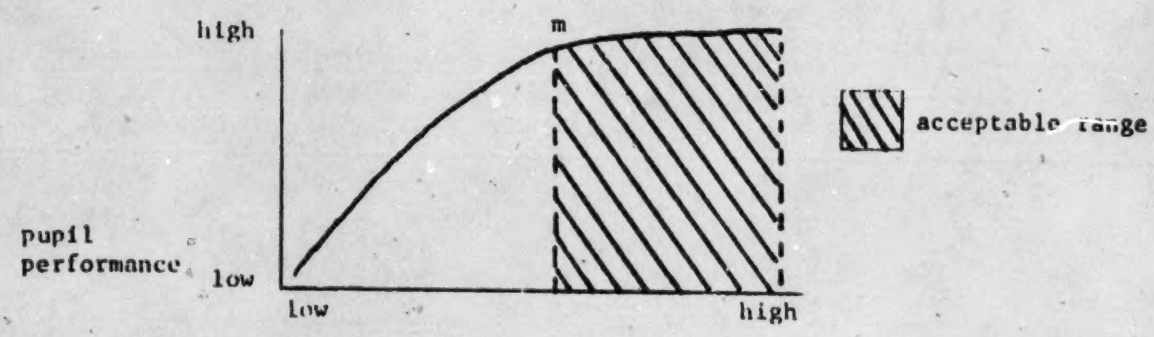
   A second type of reliability, called generalizability, considers the number of occasions on which the teacher must be observed in order for the results obtained to generalize across all occasions on which the teacher could be observed. Indices of generalizability indicate whether the number of observation occasions selected is sufficient to study each behavior and whether the behaviors trained are sufficiently stable over a reasonable number of occasions and raters to be used as correlates of pupil performance. The ultimate purpose of all classroom observation is to quantify the extent to which the teacher has implemented the behaviors taught by the training program.

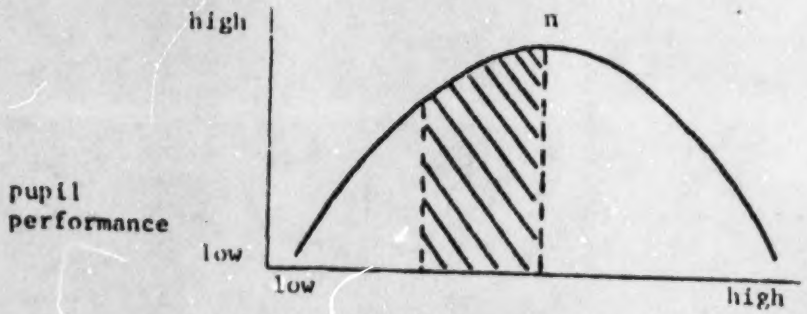
4. Measure Pre and Post Pupil Performance. As in the relative gain model, pre and post pupil performance must be measured to control for the entry level behavior of pupils. Additional variables (such as aptitude, SES, and contextual variables) that are unrelated to the instruction provided by the teacher, but which can influence pupil performance, must be taken into account.

These adjustments can be made by computing residual gain scores or by using covariance procedures. Residual gain is computed by correlating the pre- and posttest scores of all pupils, predicting a posttest score for each pupil on the basis of his or her pretest score, and subtracting this from the pupil's actual posttest score. This procedure creates a measure of gain which is independent of the pupil's initial standing and, therefore, more representative of the change due to the teacher. Analysis of covariance, which can be used to statistically control both for pretest scores and other entry level variables, represents a more efficient procedure for accomplishing the same end.

5. Plot Relationships Between Teacher Process and Pupil Product Measures. Process-product correlations are an essential element in the construction of teacher competencies. In the needs assessment and relative gain models, teacher competencies are inferred from teacher variables; only in the process-product model is the derivation of a competency empirically based. In each case, the formation of competencies should include specification of the level at which the teacher should perform a given behavior in order to be recognized as "competent." A competency is defined in terms of the level of proficiency that engenders meaningful pupil performance. The validation of various proficiency levels against meaningful classroom change is the primary contribution of the process-product model to the identification of teacher competencies.

From correlations between teacher behavior (measured by classroom observation systems), and pupil outcome (measured by criterion-referenced tests), optimal levels of teacher behavior are determined. Two possible relationships are noted below.





TEACHER BEHAVIOR B

The relationship between teacher behavior A and pupil performance reaches asymptote at point m: application of the behavior at a level greater than m nets the learner little or no improvement in criterion behavior.

Thus; for teacher behavior A, the optimal level of proficiency is m and in order to have attained the competency, the teacher must exhibit level m at the completion of training.

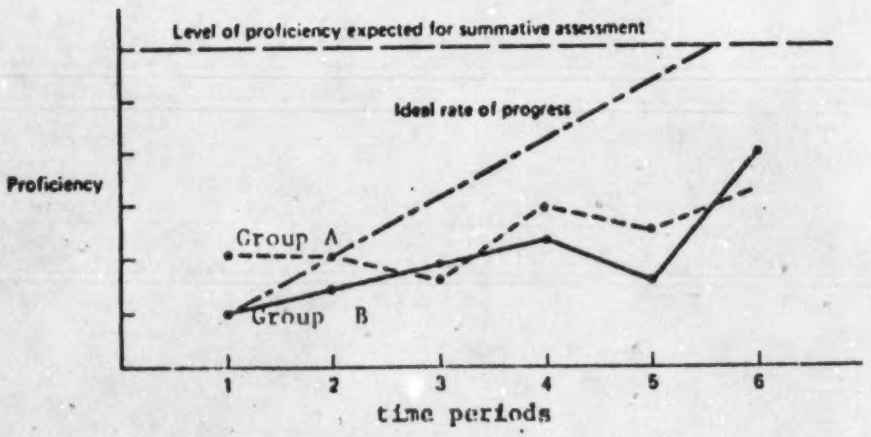
The relationship between teacher behavior B and pupil performance is markedly curvilinear, failing to reach asymptote at any point. Implementation of the behavior at a level greater than n would produce a decrement in pupil outcome. For teacher behavior B, the maximum acceptable level is n.

Relationships between process-product variables may take many other forms. But, regardless of form, they indicate to program staff which behaviors should be deleted from the training program (for lack of relationship to pupil outcome) and which levels of proficiency are most productive for each teacher behavior.

6. Construct Summary Tables. The final step in implementing the process-product model is the construction of tables summarizing the data obtained. These tables are used to compare either pre- and posttraining implementation of the behaviors taught or their use by trained and untrained teachers. They represent the most significant contribution of the process-product model to program evaluation. By indicating the extent to which the training program was able to engender the behaviors intended, and at what levels of proficiency, these tables provide the basis for judgments about the effectiveness of the total program and its components. Thus, while preceding steps investigated and confirmed process-product correlations, this step provides some indication of the program's worth. Process-product relationships, however, figure significantly in the selection of variables for which summary data are displayed. Only those variables that have shown significant relationship to pupil outcome are included on summary tables. The following examples illustrate some of the methods that can be used to display and summarize these data.

By using data from the table of frequencies and percentages, the training program can provide a continuous profile of teaching behaviors and skills for groups of teachers. A continuous profile allows the program to summarize teacher performance in a number of areas and to illustrate the precise rate of trainee improvement across observation periods for each behavior or skill. It can provide a graphic indication of the proficiency level obtained by a particular teacher. This level can then be compared with the performance of various sub-groups of teachers.

For example, profiles can be constructed to compare performance of a group of trainees on selected behaviors and skills with: (a) that of other teachers in the school district; (b) the average performance of teachers who have previously participated in the training program; or (c) an ideal performance profile, representing standards of proficiency suggested by the nature of the relationship between teacher behavior and pupil outcome, as determined in step (5) of the process-product model. A continuous profile of this type is indicated below.



Further Considerations

The process-product model is perhaps best viewed as the final component in a sequence of evaluation activities that also includes the needs assessment and relative gain models. Due partially to its quantitative nature, the process-product model may best be preceded by the relative gain or needs assessment model. Two factors suggest its terminal position in the sequence. The first concerns the "press" for immediate results which typically

surrounds the first evaluation of any program. This press makes the needs assessment and relative gain models, requiring less extensive instrumentation and little or no *in situ* observation, more attractive choices for the initial evaluation. The drain on human and fiscal resources that accompanies the process-product model may discourage relatively small training programs with limited funds and/or personal from using it, especially when the information needed may be adequately provided by the self-report responses of teachers.

Another reason for delaying process-product evaluation until after other evaluative models have been employed is the risk of negative or "null" findings. This risk is clearly greater when the method of evaluation assesses not only the effect of the training program on pupils, but also the link between this effect and specific teacher behaviors taught by that program. Programs that have already experienced some success with other more descriptive models may be more inclined to use the process-product approach and thereby risk the exposure of some shortcomings. The process-product model may be most judiciously employed when data from other sources have affirmed the efficacy of the training model and when revision of the program is a realistic option.

The amount of revision that can be tolerated and the extent to which training objectives can be operationalized can indicate the most appropriate evaluation model. If the training staff is concerned with adjusting the program to correct minor problems or to assure that the quality of training remains constant over time and setting, the decisions to be faced are considered restorative in nature. The purpose of restorative evaluation is to maintain the status quo and to provide a method of quality control. To be restorative evaluation need only specify competencies in general terms and elicit from the trainee subjective responses about their

value and attainment. Behavioral observation and pupil assessment, generally, need not be undertaken when the evaluation is intended to check for minor variations in planned outcomes.

If, however, the training staff wishes to examine larger problems in training format or content and if pupil outcome can be measured with appropriate instrumentation, the relative gain model may be appropriate. Since it produces relatively detailed pupil data, the relative gain model will likely uncover more excensive weaknesses and suggest greater revision in the program than would the needs assessment approach.

Finally, if the training staff wishes to make decisions about the rationale and assumptions underlying the program and can operationalize both teacher process and pupil outcome behaviors, the process-product model may be the method of choice. However, in this case behavioral objectives for the program must be stated in reference to both teacher and pupil, and the anticipated relationship between the two articulated. Since the process-product model examines the rationale or theory on which the training is based, it represents a more severe test of the training program than either the needs assessment or relative gain model. If the direction and magnitude of relationships predicted between teacher and pupil behavior are not confirmed, the validity of the program may be questioned. Generally, the revision indicated by process-product data is greater and more specific than that suggested by information obtained from the other two models.

Like the relative gain model, the process-product model can be expected to be viewed with some anxiety by the teacher. Not only is pupil behavior measured but also variables which may influence the behavior of pupils in spite of the teacher's performance may be present. Thus, it is hard to escape the possibility that negative findings will be ascribed to the teacher even when the contextual or antecedent conditions may make pupil growth virtually impossible. This threat may be minimized by involving

teachers in the selection of both the posttest content and the contextual and pupil variables to be statistically controlled.

Whereas the relative gain model can be employed by school districts engaged in short-term training, the process-product model seems more appropriate to the school, college, or department of education offering degree or certification programs. A training program must be of considerable duration in order to justify the time and resources required by the process-product model. Also, the combined research and evaluative functions of the process-product model often appeal to university training staff whose interests are commonly divided between research and evaluation. And, to consider a practical issue, the process-product model can be classified as research or evaluation, depending on the priorities of the funding source available. The process-product model serves these two purposes equally well by relating teacher process to pupil product, thus testing the predictions theorized, and by providing evaluative data that describe the implementation of teacher behaviors in the inservice classroom. As noted, process-product studies conducted for a specific program may actually represent a more appropriate research paradigm than program-free studies, since the former are more likely to clarify the theory or conceptual framework with which the observed process-product relationships can be explained.

The definition of evaluation implicit in the process-product model goes beyond the simple normative improvement of pupils and the subjective judgments of trainees about "what is" and "what should be," to determine the validity of the rationale or theory underlying the program. The model's capacity to test the theory employed in selecting training content and to assess pupil performance, the ultimate target of training, provides a comprehensive assessment of the program. Since most program content will derive from some rationale or theory, an evaluation of the theory is an evaluation of the program.

CONCLUSION: APPLICATIONS AND SUMMARY

The following section discusses various contexts in which the needs assessment, relative gain, and process-product models might be used. Since one of the most salient problems faced by training staff will be selecting the most appropriate model or variation thereof for a particular setting, several brief discussions are presented to illustrate the contextual characteristics most often associated with each of the models described. Key characteristics or variables that suggest the use of one model over another are identified. These descriptions necessarily include generalizations about the advantages and limitations of each model for different types of settings and therefore should be considered suggestive rather than definitive.

Case 1

A college of education has initiated a master's degree program for experienced elementary school teachers. The program is based on standard curriculum concepts and well-accepted principles of instruction. One unique aspect of the program, however, is the requirement that degree candidates take at least one year of formal course instruction at the training institution and then spend an additional year as inservice teachers applying the competencies taught. This second year takes place under the regular observation of a member of the training staff. The second year of the program is nearing completion and its first class is about to graduate when the college's administration requests evidence from the training staff that the program is fulfilling the needs of its teachers and the schools they serve. Ostensibly, the evaluation is intended to suggest program revisions that can be made before the next training cycle, but the staff suspects that a "tight money" year is causing the college to consider terminating the program. Five hundred dollars and a 1/2-time graduate student have been allocated to the evaluation for 8 weeks, at the end of which time a report is to be completed.

Suggested Model: Needs Assessment.

Distinguishing Characteristics

1. Only small adjustments in program content and structure would seem feasible. Large-scale revision or a complete reconceptualization would seem to preclude continuation of the program and, in turn, further evaluation and revision.

2. Since the training curriculum is general, it would be difficult to operationalize selected process and product variables without an extensive examination of the curriculum. It is unlikely that the required time and resources could be devoted to this end.

3. Limited time and personnel preclude classroom observation and related instrument construction, though a broad survey instrument might be created. Available funds might best be spent for questionnaire duplication and a mailing to recent graduates of the program.

4. Since evaluation results are to be used for immediate decision making rather than examination of the theory or concepts upon which the program is based, descriptive data that capture the impressions of recent graduates seem to be of most use.

5. Since there are no data affirming the efficacy of the program, it would seem some positive information is needed, integrated with suggestions for revision, for the program to have the best chance of being continued.

Case 2

A school district serving a metropolitan low-SES area has for the past six years offered a course in behavior management to its secondary school teachers. In the original proposal the program's expressed purpose was "to increase math and reading achievement through a reduction of classroom discipline problems." The course consists of 10 2-hour sessions, 1 per week for which university credit is given. Because the school district has allotted a very small proportion of the budget to instructional development, the training staff has decided to examine the format and content of their training sessions with an eye toward possible revision. Since the course is taught only during the fall, the staff has designated the final six weeks of the current training cycle and the first six weeks immediately following as the data collection period.

Suggested Model: Relative Gain (or if funds permit, Process-Product).

Distinguishing Characteristics

1. Moderate to considerable revision seems acceptable to the training staff. Objectives of the evaluation seem to involve both revision and affirmation of training content and format.

2. Due to the narrow content of the training program, i.e., behavior management, pupil and possibly teacher outcomes could be operationalized from existing program descriptions or from training materials without incurring too great an expense.

3. Time is apparently not a factor. However, personnel to train observers, to serve as observers, and to construct process instrumentation may be limited. The staff should consider the availability of student observers, the possibility of using existing rather than new process instrumentation, and the complexity (in terms of observer training) of the process behaviors to be recorded.

4. The program's expressed purpose implies a process-product link. Thus, an examination of the theory that poses this link should concern the training staff as much as a constructive evaluation of program content. The training staff might be asked to address this question in order to revise the theoretical underpinnings of the program.

5. The program has apparently been conducted for some time without criticism. Thus, its continuation seems likely even in the face of needed revision. In light of this, the training staff seems willing to accept a more stringent model.

Case 3

Four schools have been selected by the evaluation and research department of a large school district to field test a new inservice curriculum package for teaching reading in the elementary grades. This package, developed by a national research and development institute, is being implemented in the school system for the first time. Its design is purportedly based on concepts of imagery and word association as described in a recently published cognitive theory of learning. The authors of the theory claim that among the curriculum's benefits is a substantial increase in the reading comprehension of minority children who have been taught by teachers using the prescribed

methods. Since the curriculum appeared promising in a earlier product evaluation conducted by the institute that developed it, the school district has decided to fund another evaluation at a relatively high level for a one year period. The school district's primary intent is to determine whether the curriculum package can actually train teachers in the specified strategies with the effects claimed, and, hence, whether it should be disseminated to all schools in the district the following year.

Suggested Model: Process-Product.

Distinguishing Characteristics

1. Examination of the theory and major changes in the curriculum seem an acceptable outcome of the study. The school district seems concerned primarily with the effect of the curriculum on the achievement of its pupils, and this end seems best served by an examination of the relationships between teacher behavior and pupil outcome predicted by the theory.

2. Due to the explicit identification of the theory on which the program is based, intended process and product variables should be easy to operationalize by consulting the theory. Here, existing documentation may be sufficient to specify key process and product variables in terms of measurement procedures.

3. Both time and technical staff seem adequate for an extensive process-product evaluation. While some adjustments may be necessary, there probably will be no need to revert to a less comprehensive model.

4. Data will be used both to confirm the theory and to describe the program's effects on teachers. These data will also interest the research and development institute that has devised the curriculum.

5. Because the previous appraisal of the curriculum was limited to pupil outcomes, a more comprehensive evaluation is warranted if earlier findings are to be confirmed or enhanced.

Case 4

The special education department of a college of education has recently implemented a series of Saturday morning workshops designed to train inservice elementary teachers in a variety of techniques for

teaching the handicapped child in the regular classroom. The program consists of four 2-hour workshops, each conducted by a different instructor in a lecture-discussion format. The purpose of these sessions is to make regular inservice teachers aware of the different teaching strategies promoted by each of four instructors who are experts in their respective fields. The program is funded by a federal grant through the special education department. While funds have not been provided for an evaluation of these sessions, the application for renewal of program funding requires evidence supporting the "success" of the first series of workshops. The department chairman has decided to fund an evaluation in the amount of 5% of the cost of this $2,500 program.

Suggested Model: Needs Assessment.

Distinguishing Characteristics

1. Minor revisions in content and/or speakers could be made, but funds and to some extent objectives and allotted time seem to limit the program to its present format.

2. While some process variables could be specified, it is unlikely that pupil outcome could be identified at a sufficient level of detail to be attributable to the program. The primary intent of the program is to create "awareness," which seems realistic given the time and resources devoted to it. Implementation of the techniques taught might or might not occur, but in either case, it would be difficult to attribute implementation to the program per se.

3. Time and personnel permit no more than a questionnaire survey of program participants. The cost of questionnaire duplication and postage would probably just about match the amount of funds available.

4. The purpose of this evaluation is clearly descriptive. There is no implication that the workshops are linked by a theory or the techniques taught linked to pupil outcome. Specification of pupil outcome, even in broad terms, would be difficult.

5. The lack of previous evaluation data and the limited resources available for the current evaluation suggest that this initial effort should remain small and manageable, permitting only minor adjustments in workshop scope and format.

Case 5

A school district has decided to base an inservice training project on the findings of a large-scale nationwide process-product study completed the previous year. In this study, the process variable relating most significantly to the achievement of elementary pupils was the teacher's question-asking behavior (e.g., whether the question was higher order, lower order, affective, cognitive, process, or substantive). Because the school system's most pressing concern was the relatively low achievement of its secondary students, district personnel decided to develop a seven-week inservice training program to teach these questioning skills to its secondary school teachers. Since these process-product findings came from a program-free study, the staff responsible for developing training materials had to infer from these variables the nature of the materials needed. They also had to assume that such training would produce effects at the secondary school level. The project was considered somewhat risky by school district personnel since the initial process-product study had used elementary school children, and there was no guarantee that its findings would apply to older students. For this reason, the majority of funds were devoted to developing training materials, with the stipulation that additional funds would be available for more extensive evaluation and revision of materials if the initial evaluation were encouraging. Thus, the school district limited this initial evaluation to determining whether the earlier process-product findings might be generalized to the secondary school level.

Suggested Model: Relative Gain.

Distinguishing Characteristics

1. Realizing the risk involved, the school district is apparantly willing to accept either an entirely positive or an entirely negative result. Thus,

tolerance for change seems high, since the school district is willing to discontinue the program should initial results be discouraging.

2. While both process and product behaviors seem easy to operationalize, funds may be sufficient to measure only pupil outcome. An examination of pupil achievement from a group of trained teachers, vis-a-vis pupil achievement from an untrained group, could produce the result desired. That is, if the effectiveness of the program was revealed in student achievement, a later study could ascertain the exact teacher behaviors that produced the difference between trained and untrained teachers.

3. While time may not be a factor, the demand on personnel may be great, especially if criterion-referenced achievement tests are to be constructed. This can represent a considerable investment in resources, leaving little for systematic observation of the degree to which the trained teachers are implementing the questioning skills.

4. While the original process-product findings upon which the project was based must eventually be replicated in this new context, such relationships can be inferred from differences in the achievement of pupils of trained and untrained teachers. Thus, if comparison groups are used, much of the information yielded by the process-product model would be incorporated in this "control" and "experimental" version of the relative gain approach.

5. While there are no evaluative data on project materials themselves, the objectives of the study seem to require that the impact of the materials on pupils be measured. Given the findings of the original process-product study, teacher perceptions about the training program seem insufficient data upon which to decide the usefulness of subsequent and more extensive evaluations.

References

Borich, G. Implications for developing teacher competencies from process-product research. Journal of Teacher Education, Jan.-Feb. 1979, Vol, XXX, 1, 77-86.

Borich, G. and Kash, M. What the teacher effectiveness research has to say about teaching practices and student performance. Atlanta, Ga.: Georgia State Department of Education, 1979. (Also available from the Research and Development Center for Teacher Education, The University of Texas at Austin, 78712)

Medley, D.M. Teacher competence and teacher effectiveness. Washington, D.C.: AACTE, 1977.

Rosenshine, B. Review of teaching variables and student achievement. In G. Borich The Appraisal of Teaching: Concepts and Process. Reading, Mass.: Addison-Wesley, 1977.

Gage, N. L. (ed.) Mandated Evaluation of Educators. Washington, D. C.: Capitol Publications, 1973.